Environ. Sci. Technol. **2002**, *36,* 5184−5191

Probabilistic Analysis of Driving Cycle-Based Highway Vehicle Emission Factors

H. CHRISTOPHER FREY* AND JUNYU ZHENG

Department of Civil Engineering, North Carolina State University, Campus Box 7908, Raleigh, North Carolina 27695-7908

A probabilistic methodology for quantifying intervehicle variability and fleet average uncertainty in highway vehicle emission factors is developed. The methodology features the use of empirical distributions of emissions measurement data to characterize variability and the use of bootstrap simulation to characterize uncertainty. For the base emission rate as a function of mileage accumulation under standard conditions, a regression-based approach was employed in which the residual error terms were included in the probabilistic analysis. Probabilistic correction factors for different driving cycles, ambient temperature, and fuel Reid vapor pressure (RVP) were developed without interpolation or extrapolation of available data. The method was demonstrated for tailpipe carbon monoxide, hydrocarbon, and nitrogen oxides emissions for a selected light-duty gasoline vehicle technology. Intervehicle variability in emissions was found to span typically 2 or 3 orders of magnitude. The uncertainty in the fleet average emission factor was as low as ±10% for a 95% probability range, in the case of standard conditions, to as much as −90% to +280% when correction factors for alternative driving cycles, temperature, and RVP are applied. The implications of the results for method selection and for decision making are addressed.

Introduction

The National Research Council (NRC) recommends that efforts be conducted to quantify uncertainties in highway emission estimates (*1*). Such estimates are widely used at the state and federal level for regulatory, planning, and other decision-making purposes involving substantial resources (*2*). Thus, there is incentive to make air quality management decisions that are robust to uncertainty.

Kini and Frey (*3*) and Pollack et al. (*4*) have reported results for probabilistic analysis pertaining to aspects of the U.S. Environmental Protection Agency's (EPA's) Mobile5b and the California Air Resources Board EMFAC7F emission factor models, respectively. Both studies focused on a bottoms-up approach to assessing uncertainty in emission factors based upon statistical analysis of emission test data used to develop the model. Others (*5*) used a bootstrap approach to calculate confidence intervals for the speed correction factor in Mobile5a but retained the functional form of a curve fit employed by U.S. EPA in their analysis.

* Corresponding author. E-mail: frey@eos.ncsu.edu. Tel: (919) 515-1155. Fax: (919) 515-7908.

Compared to the Kini and Frey study, this paper introduces additional methodological tools required to deal with correction factors for which only relatively small data sets are available, with case study examples for temperature and Reid vapor pressure (RVP) corrections. To demonstrate the method and insights obtained from it, detailed estimates of uncertainty are provided for a light-duty gasoline vehicle (LDGV) technology group for carbon monoxide, nitrogen oxides (NO_x), and hydrocarbon (HC) emission factors for each of 11 driving cycles at standard temperature and RVP and for situations in which emissions are corrected for other values of temperature and RVP.

The case study for this paper is based upon the Mobile5b emission factor model (*6*−*9*). A new emission factor model, Mobile6, was released after this work was completed (*10*). Insufficient data and documentation were available regarding Mobile6 during the time frame of this study from which to develop a comparative probabilistic analysis. While Mobile6 uses improved data and is expected to provide more representative (accurate) emission estimates than Mobile5 (e.g., Mobile6 includes facility-specific driving cycles, which Mobile5 does not), both models share similar approaches regarding the use of multiplicative correction factors to adjust a base emission rate to nonstandard conditions. Therefore, the *methodological* issues regarding uncertainty analysis are similar for both models. The application of the methods presented here for quantification of variability and uncertainty of driving cycle-based highway vehicle emission factors is primarily with respect to Mobile5. However, to illustrate that the results are similar when applied to either Mobile5 or Mobile6, an analysis of uncertainty in the mean emissions for the driving cycles used to develop the Mobile6 speed correction factor is included for comparison with analogous results for Mobile5.

Sources of Variability and Uncertainty. Variability refers to the heterogeneity across different elements of a population over time or space. Uncertainty is a lack of knowledge about the true value of a quantity. Uncertainty in emissions is typically attributable to the following: (1) random measurement errors (lack of precision); (2) systematic errors (bias or lack of "accuracy") such as would be caused by imprecise calibration or use of surrogate data (e.g., laboratory tests of vehicles rather than on-road measurements); (3) lack of empirical basis such as would occur when measurements have not been taken or when estimating emissions for a future source; and (4) human error such as random mistakes in entering or processing data. Variability can be represented by a frequency distribution. Uncertainty can be quantified as a probability distribution (*12*−*19*).

Variability and Uncertainty in Highway Vehicle Emission Factors. Emissions vary from one vehicle to another because of differences in design, operation, maintenance, and fuel composition. Emissions measurements using specific driving cycles attempt to control for operation by imposing a specific speed versus time profile and for fuel composition. Several researchers describe the inherent variability in emissions measurements obtained using a variety of testing methods (*20*−*23*). The main focus of the uncertainty analysis is on characterizing random and systematic errors associated with estimates of fleet average emissions. Random errors are characterized based upon statistical analysis of random sampling error. Systematic errors are characterized based upon deviations of the point estimate predictions of the model when compared to mean values inferred directly from available data.

10.1021/es0114308 CCC: $22.00
Published on Web 10/30/2002

Modeling Assumptions and Input Data

The methodological approach includes the following elements: (1) development of a simplified empirical emission factor model, similar to that of Mobile5b; (2) collection of emission test data for an example case study; and (3) probabilistic analysis and modeling techniques. The first two are described here, and the third is presented with the case study results.

Brief Review of the Mobile5b Model. Mobile5b estimates emission factors for CO, HC, and NO_x by calculating a base emission rate (BER) for a standard driving cycle and standard conditions (e.g., ambient temperature, RVP) associated with a given mileage accumulation (odometer reading). The BER is adjusted to other conditions, such as different driving cycles, ambient temperature, and RVP, using correction factors (*7*−*10*). The BER is developed separately for different vehicle types (e.g., light-duty gasoline vehicles) based upon an assumed mix of technology groups, the latter of which are typically characterized on the basis of fuel delivery and emission control systems (e.g., throttle body injected engines, three-way catalysts). Emission control systems are assumed to undergo "deterioration" as a function of mileage accumulation. Curve fits are used in Mobile5b for the BER and for each correction factor. The curve fits are typically based upon regression analysis of driving cycle data. Mobile5b determines point estimates for each step in the calculation process.

Simplified Probabilistic Emission Factor Model. The BER in Mobile5b is intended to represent emissions for bag 2 of the FTP driving cycle, which is taken as the reference point to which correction factors are applied. However, to obtain a large data set representative of the on-road vehicle fleet, EPA used inspection and maintenance program data obtained using the IM240 test procedure. The IM240 test is based upon a portion of the speed profile used in the FTP. EPA developed a regression equation for each of CO, HC, and NO_x emissions to convert the IM240 measurements to an equivalent FTP emission estimate. EPA used logarithmic transformations to develop the IM240 to FTP regression equations for CO and HC and used a linear formulation in the case of NO_x. EPA did not account for the residual error term of the regression equations, which reflects the intervehicle variability in emissions that is not explained by the model. The residual error term is multiplicative for CO and HC, and it is additive for NO_x, because of the formulations assumed. Here, the residual errors were characterized as empirical distributions based upon analysis of the data sets used by EPA to develop the IM240 to FTP regression models.

The estimated FTP emissions were used by EPA to develop a linear regression equation for emissions versus mileage accumulation for each of the three pollutants, thereby introducing a second residual error term. However, the residual error term was not normally distributed, which violates the assumption of least-squares regression. The residuals for the logarithm of emissions were more nearly normally distributed. Therefore, a log-linear regression is used here instead, which differs from the approach used by EPA. For CO and HC, the BER equation is

$$\mathrm{BER} = \exp\{\mathrm{ZML} + \mathrm{DR} \times \mathrm{MA} + \epsilon_1\}\epsilon_2 \qquad (1)$$

For NO_x emissions, the BER equation is

$$\mathrm{BER} = \exp\{\mathrm{ZML} + \mathrm{DR} \times \mathrm{MA} + \epsilon_1\} + \epsilon_2 \qquad (2)$$

where BER is the base emission rate (g/mi), ZML is the zero mile level emission constant (logarithm of g/mi), DR is the deterioration rate constant for mileage less than or equal to 50 000 mi (logarithm of g/mi^2), MA is the mileage accumulation less than or equal to 50 000 mi (mi), ϵ_1 is the residual error distribution for the BER regression equation (logarithm of g/mi), and ϵ_2 is the residual error distribution for the IM240 to FTP regression equation (dimensionless ratio for CO and HC, g/mi for NO).

The BER represents emissions for the FTP driving cycle under standard conditions, including ambient temperature of 75 °F and fuel RVP of 9.

The emission factor for nonstandard conditions is estimated using multiplicative correction factors, each of which is a dimensionless ratio. Three correction factors are evaluated empirically based upon data analysis: (1) speed correction factor (SCF); (2) temperature correction factor (TCF); and (3) RVP correction factor (RVPCF):

$$\mathrm{EF} = \mathrm{BER} \times \mathrm{SCF} \times \mathrm{TCF} \times \mathrm{RVPCF} \qquad (3)$$

The SCF is the ratio of emissions on a non-FTP driving cycle to the emissions on the FTP driving cycle. The TCF is the ratio of emissions at ambient temperature T on an FTP test to the emissions on the standard FTP test, which has a temperature of 75 °F. The RVPCF is the ratio of emissions for a nonstandard RVP to that of the standard RVP of 9.0, with both evaluated using the FTP test.

The functional form of eq 3 is similar to but not the same as that in Mobile5b. The Mobile5b model employs curve fits for correction factors, and the RVPCF curve fit includes temperature as an explanatory variable. In the approach used here, curve fits are not employed to avoid introduction of systematic errors associated with any particular model formulation.

Collection of Emission Test Data. Data for the case study are based upon LDGV Technology Group 8, which has a throttle body fuel injection system and a three-way catalyst. The data used for the SCF analysis involved measurements of multiple vehicles on multiple driving cycle tests. The tests include bag 2 of the FTP, as well as the LSP1, LSP2, LSP3, NYCC, SCC12, SCCC36, HFET, HSP1, HSP2, and HSP3 test procedures. Each procedure is characterized by a different speed trace. The average speeds vary from 2.5 mph for LSP1 to 64 mph for HSP3. A set of 35 vehicles was tested on each of the NYCC, SCC12, FTP bag 2, SCC36, and HFET procedures. Fourteen of the vehicles were also tested on the LSP1, LSP2, and LSP3 cycles. Eight of the vehicles were also tested on the HSP2 and HSP3 cycles, while four were tested on the HSP1 cycle. The ratio of each vehicle's emissions on a nonstandard cycle to its emissions on the FTP bag 2 cycle were calculated, and an empirical distribution of the intervehicle variability in the ratio was developed. The uncertainty in the average SCF was characterized based upon the sampling distribution of the mean, which is influenced both by the sample size and by variability.

The available data sets for estimating variability and uncertainty in the TCF and RVPCF are much smaller than for the SCF. For example, for a selected set of typically only three or four vehicles, several repeated FTP tests were run at the standard temperature of 75 °F, and then several repeated FTP tests were run at a different temperature, such as 50 °F.

In developing empirical correction factors, extrapolation of the actual test data is avoided by considering conditions only for which test data are available. These include temperatures of both 75 and 50 °F at RVP = 9 psi and a temperature of 50 °F at RVP = 13 psi. Thus, a temperature correction factor is first applied to represent the conditions of lower temperature, and then an RVP correction factor is applied to represent conditions of high RVP at the lower temperature.

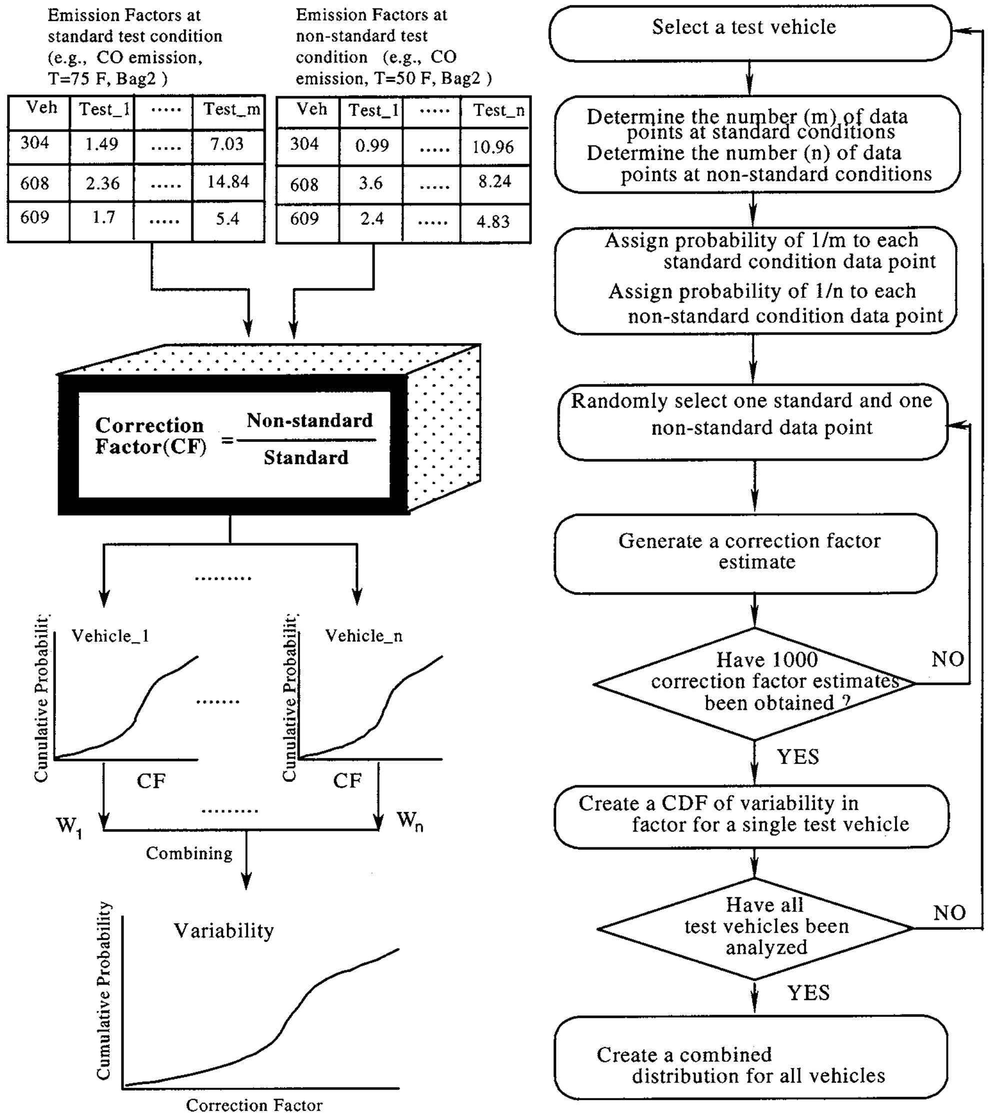

FIGURE 1. Simplified flow diagram of method for estimating variability in correction factors.

Quantification of Intervehicle Variability in Correction Factors

Intervehicle variability in correction factors was estimated by sampling from the available emissions measurement data to construct cumulative probability distribution functions. There is intravehicle variability reflected by differences in the replicate measurements for a given vehicle, and there is intervehicle variability reflected by differences in average measurements when vehicles are compared. A distribution for variability for a single vehicle was developed by analyzing the replicate measurements for that vehicle. A combined distribution of variability among the three or four vehicles tested was developed by combining their individual distributions into a single mixture distribution. The observed intravehicle variability could also be interpreted as variability in emissions associated with differences in operation, which is a factor that contributes to intervehicle variability. Even though vehicles are tested with respect to a standard speed profile, the test driver is allowed to deviate from the speed trace within a tolerance, and such deviations can lead to variability in emissions (*20*). Thus, the combined mixture distribution is interpreted as an indication of overall intervehicle variability.

The development of the mixture distribution includes the following tasks: (1) simulation of a distribution of variability for an individual vehicle; (2) development of weighting factors for each vehicle; and (3) simulation of a mixture distribution including all available vehicles. The method is summarized in Figure 1. More details are available elsewhere (*19*). It was assumed that each vehicle tested is equally representative of the on-road fleet. Therefore, if the number of vehicles tested was v, the weight assigned to each vehicle when the mixture distribution was developed is $1/v$.

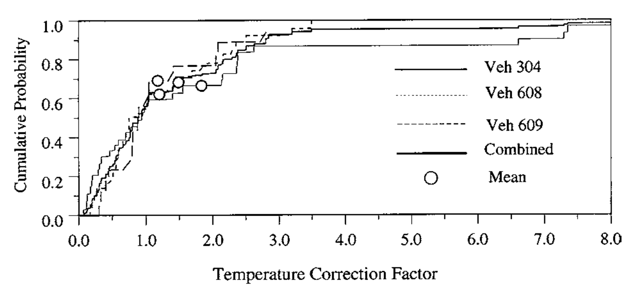

FIGURE 2. Estimated intervehicle variability in temperature correction factor for CO technology group 8.

As an example, the individual and mixture CO TCF distributions for three vehicles are shown in Figure 2. The stepwise nature of the empirical cumulative distribution functions (CDFs) reflects the finite number of possible combinations of correction factors if m and n are approximately 3 or 4 each. The lower tail of the combined (mixture) distribution asymptotically approaches the lower tail of the distribution that has the lowest values of TCF, and the upper tail of the mixture distribution asymptotically approaches the upper tail of the distribution that has the highest values. In this case, vehicles 608 and 609 have similar variability, whereas vehicle 304 has a wider range of variability than the other two. The correction factor varies from approximately 0.13 to 7.3 over a 95% probability range, which is a span of more than 1 order of magnitude. The mean value is 1.43. This implies that, on average, emissions of CO are expected to increase by 43% if temperature decreases from 75 to 50 °F.

For the CO RVPCF, the average value varies from approximately 1.49 to 4.85 among the three vehicles, with a weighted average of 2.5. The range of intervehicle variability is from approximately 0.23 to a value of 9.8 over a 95% probability range, which is a range of more than 1 order of magnitude. For all three pollutants, there is a possibility that emissions may be lower, but on average, it is expected that emissions will be higher if a higher RVP fuel is used. The simulated mean correction factors were compared with an analytical calculation of the mean values (explained in the Supporting Information), and the simulated means were shown to be unbiased. The distributions of variability in the correction factors are summarized in Tables S-1, S-3, and S-5 (in Supporting Information) for CO, HC, and NO_x, respectively.

Quantification of Uncertainty in Mean Correction Factors

When emission inventories for motor vehicles are being developed, average emission factors for the on-road fleet are more useful than emission rates for individual vehicles. The mean is a statistic calculated from a random sample of data; therefore, it is a random variable. Under idealized conditions, the sampling distribution for the mean can be approximated with a normal distribution if the sample size is sufficiently large, if there is a sufficiently small range of variability in the data, or both. However, the data sets used by EPA for developing correction factors are small, and there is a high degree of variability and positive skewness in the data sets. Therefore, it is not reasonable to assume normality for the sampling distribution of the mean. Instead, the numerical technique of bootstrap simulation is employed to estimate the sampling distribution of the mean. Bootstrap simulation was introduced by Efron in 1979 for the purpose of estimating confidence intervals for statistics (*13*). Bootstrap simulation does not require any assumptions regarding the shape of the sampling distribution.

The version of bootstrap simulation employed here involves randomly simulating, with replacement, a data set of the same sample size as the original data set to create a bootstrap sample, which is a randomized version of the original data set. For each bootstrap sample, a replicate of the statistic of interest (e.g., mean) is calculated. The process is repeated many times to obtain multiple randomized estimates of the statistic. Typically 200 bootstrap replications are sufficient to estimate confidence intervals (*13*). However, since the intent here is to estimate the sampling distribution, 1000 replications are used. The method for simulating uncertainty in the correction factors is summarized in Figure 3 and is explained in more detail elsewhere (*19*). So as not to underestimate the uncertainty associated with random sampling error, we let the sample size of each bootstrap sample be the minimum of the number of measurements at the standard condition or at the non-standard condition.

Figure 4 illustrates the results of analysis of the mean values of the CO TCF, showing the individual distributions for uncertainty in the mean of each of three vehicles and the equally weighted mixture of all three. The range of uncertainty in the average correction factor is from approximately 0.38 to 3.5 over a 95% probability range. Thus, the range of uncertainty in the mean is less than the range of intervehicle variability. However, because only three vehicles were used in the testing, and because only a small number of tests were conducted at each temperature (typically only three or four), the range of uncertainty in the mean is substantial and spans a factor of ~10. The distributions of uncertainty in the correction factors are summarized in Tables S-2, S-4, and S-6 (in Supporting Information) for CO, HC, and NO_x, respectively.

Quantification of Variability and Uncertainty in the Emission Factors

The emission factors for CO and HC are calculated using eqs 1 and 3, and for NO_x, they are calculated using eqs 2 and 3. To estimate the intervehicle variability in emission factors, distributions for intervehicle variability are used for the residual errors and correction factors as summarized in Tables S-1, S-3, and S-5 (in Supporting Information) for CO, NO_x, and HC, respectively. A software package, Analytica, was used to simulate variability in the tailpipe emission factors using Monte Carlo simulation.

Fleet average uncertainty in LDGV tailpipe emission factors is calculated in a manner similar to that for intervehicle variability, except that sampling distributions for mean values are used instead of frequency distributions for intervehicle variability. The distributions for fleet average uncertainty used for the inputs to the emission factor model are summarized in Tables S-2, S-4, and S-6 (in Supporting Information) for CO, NO_x, and HC, respectively.

For each of the three pollutants, three sets of results were developed, representing different combinations of probabilistic assumptions. All three include probabilistic assumptions for the base emission rate. One set is based upon the use of only the speed correction factor. The second set is based upon adjustment of emissions for an ambient temperature of 50 °F, using the probabilistic temperature correction factor. The third set is based upon additional adjustment of the emission factors for a fuel RVP of 13, using the probabilistic RVP correction factor. These three sets of results are presented for both intervehicle variability in emissions and fleet average uncertainty in highway vehicle CO emissions in Tables 1 and 2, respectively. Results for intervehicle variability and fleet average uncertainty are given for HC in Tables S-7 and S-8, respectively, and for NO_x in Tables S-9 and S-10, respectively (in Supporting Information).

Intervehicle Variability in Emission Factors. Table 1 contains CO emission factor estimates for 11 driving cycles. The reported point estimate is obtained based upon the methods and assumptions of Kini and Frey (*3*). Selected results are discussed here to illustrate the types of findings obtained from the probabilistic analysis. For example, for the low average speed LSP1 cycle, the 95% probability range of intervehicle variability is from 1.27 to 241 g/mi when variability is considered only in the BER and SCF. This range is more than 2 orders of magnitude. The point estimate is based upon an analysis in which the skewness of the residual error term of the IM240-to-FTP regression model is not considered, which is similar to the approach used in the Mobile5b emission factor model. Furthermore, the point estimate is based upon a curve fit used by EPA for the speed correction factor. There are biases in the point estimate based upon the use of a speed correction factor curve fit and failure to properly account for residual errors in the Mobile5b emission factor model.

When the temperature correction factor is applied, the estimated range of variability increases. For CO emissions

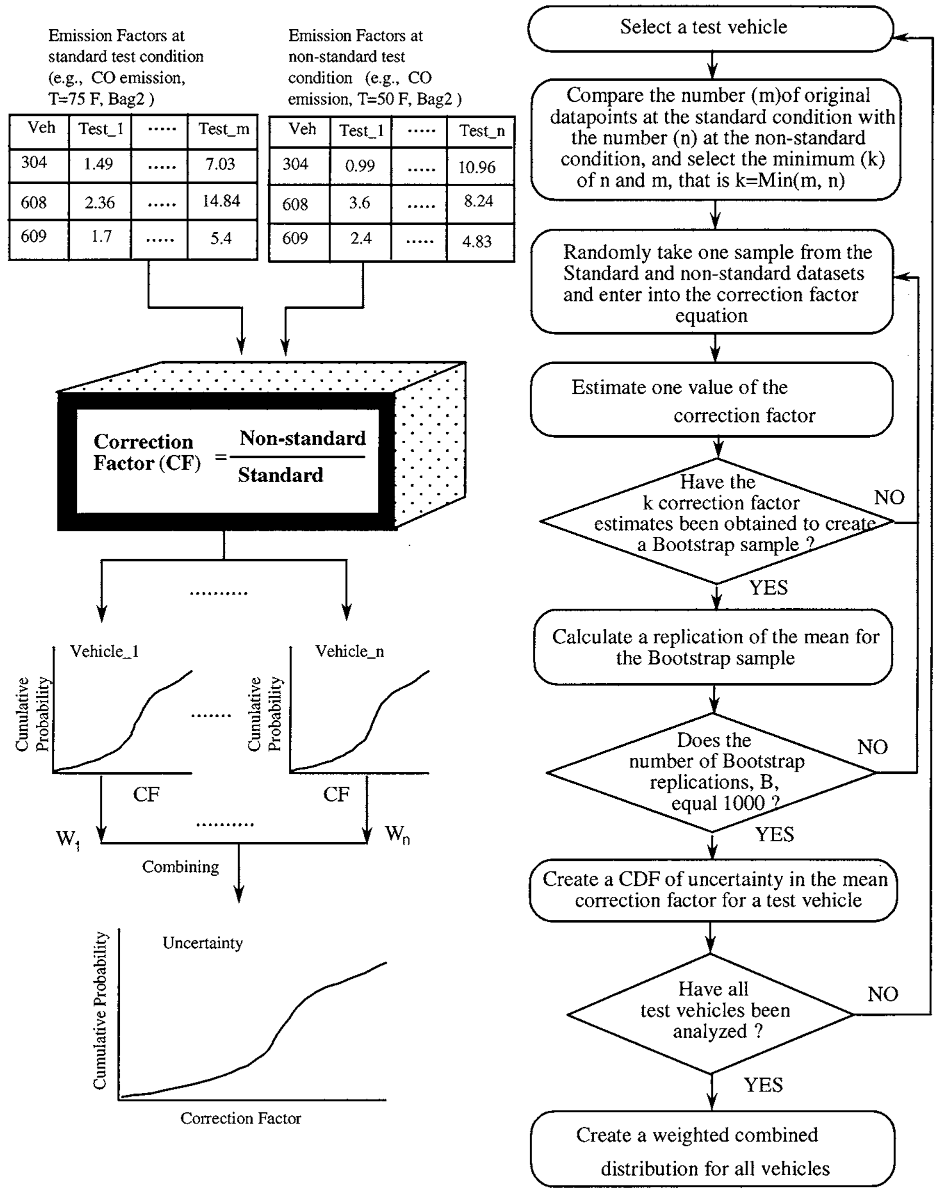

FIGURE 3. Simplified flow diagram of method for estimating uncertainty in correction factors.

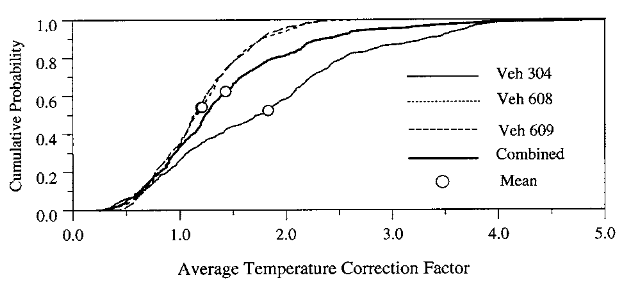

FIGURE 4. Estimated fleet average in temperature correction factor For CO technology group 8.

on the LSP1 cycle, the predicted 95% probability range for variability is from 0.56 to 356 g/mi, which is substantially wider than the range of variability at the standard ambient temperature. When the emissions estimate is adjusted for a fuel RVP of 13 psi instead of 9 psi, the 95% range of variability increases to an interval from 0.52 to 1280 g/mi.

Similar trends regarding the increase in estimated inter-vehicle variability resulting from the application of additional correction factors are observed for the other driving cycle emission estimates for CO tailpipe emissions. These trends are also observed for HC and NO_x emission factor estimates. The effects of variability in TCF and RVPCF on variability in emission factors are substantial.

Uncertainty in Mean Emission Factors. Probabilistic estimates of fleet average uncertainty in CO emission factors are summarized in Table 2. Systematic and random errors are reported. The systematic error is the point estimate minus the mean. In general, the mean values tend to be higher than the point estimates obtained using the deterministic modeling methodology as employed by Mobile5b. For CO and HC, the residual error distribution, ϵ_2, has a mean of greater than 1, implying that the Mobile5b model is systematically underestimating the average emission rate because the residual error was not properly accounted for. Furthermore, although many of the input distributions for the uncertainty analysis are symmetric, a multiplicative model will typically yield positively skewed distributions for the product.

The random error is described in terms of how the 95% confidence interval compares to the mean on a relative basis. When uncertainty is relatively small, the random error is approximately symmetric and can be described as a "plus or minus" range. For example, the 95% probability range of uncertainty in the mean CO emission factor for LSP1 without

TABLE 1. Characterization of Intervehicle Variability in Estimated Tailpipe CO Emission Factors for Technology Group 8[a]

		$T = 75$ °F, RVP = 9 (only SCF)				$T = 50$ °F, RVP=9 (SCF and TCF)				$T = 50$ °F, RVP=13 (SCF, TCF, and RVPCF)			
driving cycle	speed (mph)	point estimate	2.5th percentile	mean	97.5th percentile	point estimate	2.5th percentile	mean	97.5th percentile	point estimate	2.5th percentile	mean	97.5th percentile
LSP1	2.45	57.2	1.27	44.0	241	81.7	0.627	58.2	377	223	0.551	150	1040
LSP2	3.64	38.1	1.77	42.5	247	54.4	0.753	66.3	419	149	0.703	164	1290
LSP3	4.02	34.3	1.24	54.2	301	49.0	0.644	77.7	592	134	0.552	218	1470
NYCC	7.10	18.8	2.27	37.9	193	26.9	0.974	65.2	398	73.3	0.968	172	1060
SCC12	12.1	10.4	0.75	16.0	82.1	14.9	0.435	24.2	130	40.6	0.353	69.0	442
FTP BAG2	16.1	7.46	0.80	10.5	46.4	10.7	0.343	16.1	96.7	29.1	0.287	43.4	305
SCC36	35.9	5.87	0.40	8.16	44.3	8.39	0.218	11.5	81.9	22.9	0.173	30.8	212
HFET	48.4	4.06	0.31	4.74	25.6	5.80	0.143	7.48	51.4	15.8	0.106	19.8	137
HSP1	50.9		0.36	5.06	22.4		0.143	7.91	55.2		0.124	20.9	148
HSP2	57.6		0.02	0.29	1.38		0.00806	0.436	3.17		0.00642	1.15	8.53
HSP3	64.3		0.01	0.29	1.34		0.00729	0.479	3.10		0.00638	1.16	7.77

[a] Emission factors are in grams per mile. Point estimate is a deterministic estimate of the emission factor obtained as described by Kini and Frey (1997). The 2.5th and 97.5th percentiles describe a 95% probability range for the emission factor. The mean emission factors were obtained from probabilistic simulation.

TABLE 2. Characterization of Fleet Average Uncertainty in Estimated Tailpipe CO Emission Factors for Technology Group 8[a]

		$T = 75$ °F, RVP = 9 (only SCF)						$T = 50$ °F, RVP = 9 (SCF and TCF)						$T = 50$ °F, RVP = 13 (SCF, TCF, and RVPCF)					
						random error (%)						random error (%)						random error (%)	
driving cycle	speed (mph)	2.5th %-ile	mean	97.5th %-ile	systematic error	(−)	(+)	2.5th %-ile	mean	97.5th %-ile	systematic error	(−)	(+)	2.5th %-ile	mean	97.5th %-ile	systematic error	(−)	(+)
LSP1	2.45	15.3	36.9	59.8	23.6	−59	62	12.5	53.4	156	28.3	−77	192	15.1	148	567	75	−90	282
LSP2	3.64	11.7	38.9	64.5	0.92	−70	66	9.09	55.6	159	−1.2	−84	187	14.2	154	583	−5	−91	279
LSP3	4.02	19.4	50.9	84.1	−16.4	−62	65	14.5	73.4	227	−24.4	−80	209	20.0	222	727	−88	−91	227
NYCC	7.10	24.9	33.1	40.8	−11.8	−25	23	13.5	47.2	124	−20.3	−71	163	16.1	130	464	−57	−88	256
SCC12	12.1	10.3	14.3	18.2	−2.89	−28	27	5.81	20.4	55.4	−5.5	−72	171	6.49	56.8	187	−16	−89	229
FTP BAG2	16.1	7.91	8.78	9.66	−0.71	−10	10	3.57	12.6	33.7	−1.9	−72	168	4.40	34.7	112	−6	−87	224
SCC36	35.9	5.30	6.65	8.15	−0.33	−20	23	2.71	9.52	25.3	−1.13	−72	165	3.10	26.3	88.4	−3	−88	236
HFET	48.4	3.27	4.16	5.08	0.21	−21	22	1.72	5.95	15.7	−0.15	−71	165	1.98	16.5	57.1	−1	−88	246
HSP1	50.9	2.80	4.50	6.44		−38	43	1.72	6.40	16.7		−73	161	2.12	18.2	65.7		−88	262
HSP2	57.6	0.15	0.26	0.37		−42	43	0.10	0.37	0.99		−72	170	0.12	1.03	3.46		−88	236
HSP3	64.3	0.12	0.27	0.42		−57	57	0.08	0.38	1.09		−79	183	0.11	1.04	3.85		−90	272

[a] Emission factors are in grams per mile. Point estimate is a deterministic estimate of the emission factor obtained as described by Kini and Frey (1997). Systematic error = point estimate -mean. The 2.5th and 97.5th percentiles describe a 95% probability range for the emission factor. The mean emission factors were obtained from probabilistic simulation. Random error (−) = (2.5th percentile-mean)/mean × 100. (+) = (97.5th percentile − mean)/mean.

any additional corrections is −59% to +62%, or ∼±60%. However, when uncertainty is relatively large, the random error becomes asymmetric. For example, for the LSP1 CO emission factor estimate with both TCF and RVPCF, the range of uncertainty is −90% to +282%. The asymmetric uncertainty range results from the fact that emission factors cannot be negative. Therefore, the sampling distribution of the mean is positively skewed when the range of uncertainty is large relative to the mean value.

For all driving cycles, the range of uncertainty in the mean CO emission factors with all correction factors applied is approximately −90% to +240% in many cases. For HC, the range of uncertainty in the mean emission factors with all correction factors applied is approximately −90% to +220% in most driving cycles. The range of uncertainty in the NO_x emission factors is not quite as large, ranging from approximately −60% to +120%. In all cases, the range of uncertainty in the emission factor with both TCF and RVPCF is substantially larger than for the base case conditions of ambient temperature and fuel RVP.

Identifying Key Sources of Uncertainty

Tables S-11, S-12, and S-13 (in Supporting Information) show the sample correlations for the uncertain emission factors for CO, HC, and NO_x, respectively, calculated for each driving cycle when the SCF, TCF and RVPCF are all applied with respect to uncertainty in each individual model input. For all 11 driving cycles and for all three pollutants, the largest sample correlation coefficients are asssociated with the input uncertainty assumptions for TCF and RVPCF, with values of approximately 0.5−0.75 in most cases. In contrast, the uncertainty in the residual error terms, ϵ_1 and ϵ_2, typically have sample correlations of less than 0.1 in magnitude for CO and HC and less than 0.3 in magnitude for NO_x. The uncertainty in the SCF also contributes only modestly to overall uncertainty in the emission factors as reflected by a sample correlation of approximately 0.1−0.3 in most cases. There are some exceptions to these general trends. For example, for NO_x emissions, the uncertainty in the SCF contributes more to the range of uncertainty in the case of the HSP3 driving cycle than does any other input.

Comparison of Uncertainty Estimates for Mobile5 and Mobile6

It was not possible to fully develop an uncertainty analysis of Mobile6 similar to that for Mobile5 within the scope of resources available for this work. Specifically, it was not possible to obtain complete data sets in order to evaluate uncertainty in the base emission rates for Mobile6. However, EPA has reported statistical summaries of average emissions, the standard deviation of emissions, and the sample size for 15 driving cycles that were used to estimate emissions for tier 1 vehicles as a basis for developing speed correction factors. Based upon these summaries, uncertainty in the average emissions for each driving cycle was estimated. The results are shown in Figures S-1, S-2, and S-3 (in Supporting Information) for HC, CO, and NO_x, respectively.

For the analysis of uncertainty in average emissions for Mobile5 including the speed correction factors, the average range of uncertainty in the mean for 11 driving cycles was approximately ±31% for HC, ±40% for CO, and ±45% for NO_x. For the 15 driving cycles that were used in developing the Mobile6 speed correction factor, the average range of uncertainty in the mean was approximately ±35% for HC, ±46% for CO, and ±38% for NO_x. The results for the relative ranges of uncertainty are similar for Mobile5 and Mobile6 even though they are based upon a different sample of vehicles and even though few of the cycles are common to both models. Another similarity is that the analyses indicate that some of the driving cycles are redundant with each other. For example, the LSP1, LSP2, and LSP3 cycles of Mobile5 produce statistically similar emission rates for all three pollutants. The ART-AB and the FWY-D, -E, -F, and -G cycles of Mobile6 produce statistically similar emissions for all three pollutants. For example, the 95% confidence intervals for these five cycles overlap substantially, including a common range of 0.028−0.043 g/mi for HC, 0.86−1.80 g/mi for CO, and 0.17−0.25 g/mi for NO_x. Thus, the methodology demonstrated for uncertainty analysis of Mobile5 would likely lead to comparable results and similar insights when applied more extensively to Mobile6.

Results and Discussion

In this work, all input variability distributions to the emission factor models were characterized based upon empirical distributions, rather than based upon assumed parametric probability distributions (e.g., normal, log-normal). Thus, the analysis was without additional assumptions regarding the shape of the probability distributions and without introducing biases associated with curve fits. For the uncertainty analysis, normal distributions were used only when justified by the central limit theorem. For the TCF and RVPCF, empirical distributions based upon bootstrap simulation were used to characterize uncertainty in mean values.

The range of both variability and uncertainty in the TCF and RVPCF was generally large and contributed more to the range of variability and uncertainty in emission factors than did other model inputs. Additional data would lead to more precise estimates of intervehicle variability and would typically lead to narrower ranges of estimated uncertainty.

Analysts and decision makers are typically interested in predicting average emissions for fleets of vehicles rather than in knowing emission rates for individual vehicles. In many cases, the range of uncertainty in the average emissions is so large that simplifying assumptions based upon normality cannot be employed. For example, some emission factors were found to have uncertainty ranges of −80% to +220% of the mean value. The asymmetry reflects the fact that the emission factors are nonnegative quantities and is influenced by both the large intervehicle variability in emissions and the relatively small sample sizes of data sets from which the emission factors were developed.

The levels of uncertainty in the emission factors can be evaluated in terms of data quality objectives, as recommended by the National Research Council. Is the range of uncertainty acceptable? If not, what can be done to reduce uncertainty? One approach for reducing uncertainty is to collect more data for those specific inputs to the model that contribute most to uncertainty in the emission factor. Uncertainty in the TCF and RVPCF was typically the dominant source of uncertainty in the estimated emission factors. In contrast, uncertainty in the speed correction factor and regarding the residual errors of regression equations used in the model typically contributed far less to overall uncertainty. Therefore, it would be most beneficial to prioritize data collection on the TCF and RVPCF. It should be noted, however, that not all sources of uncertainty are quantifiable, especially those that might lead to systematic error compared to true on-road emissions. Thus, a qualitative consideration in uncertainty analysis is whether data are representative of the system being modeled and, if not, whether more representative data can be obtained as a basis for future model refinement.

The analysis focused on one LDGV technology group. It is only speculative to generalize these results to other technology groups. However, the selected group represents a substantial portion of on-road vehicles and has relatively large data sets compared to other technology groups. Therefore, it is likely that the quantitative uncertainty estimates would be wider for some though perhaps not all other technology groups. Similar insights regarding uncertainty estimates were obtained based upon analysis of driving cycles used to develop the speed correction factor in Mobile6.

Comparison of uncertainty estimates for individual driving cycles reveals that in some cases two or more driving cycles produce statistically similar results and, therefore, are redundant. The identification of redundant cycles, such as the LSP1 and LSP2 cycles, or the ART-AB and FWY-D, -E, -F, and -G cycles, presents opportunities to reduce the cost of data collection by focusing upon a smaller set of nonredundant cycles.

Although the ranges of uncertainty in the emission factors can be large, this does not imply that the emission factors are meaningless. The significance of the range of uncertainty is context-dependent. Aside from using uncertainty analysis as a tool for prioritizing additional data collection, uncertainty analysis can be applied to emission inventories. A quantitative assessment can be made of the likelihood with which an emission budget will be met. A decision maker can use this information to make tradeoffs between emissions management strategies and the confidence with which the budget will be met. Furthermore, probabilistic emission inventories can be used as input to air quality models to determine the likelihood that ambient air quality management goals will be achieved and to develop strategies that produce an acceptable confidence level of air quality benefits.

The methodology for uncertainty analysis demonstrated here can be applied in future work to facilitate comparisons of Mobile5 with other models. For example, a difference in predictions of two models of only 20%, or even a factor of 2 or more in some cases, may not be statistically significantly different. However, for situations in which two models give statistically significantly different predictions, a qualitative evaluation and choice can be made as to which model is more representative of the system being modeled. For example, future comparisons of Mobile6 with Mobile5b would be informed by the quantitative uncertainty analysis of Mobile5b presented here coupled with an uncertainty analysis of Mobile6. Furthermore, an assessment of key sources of uncertainty in Mobile6 would help identify priorities for improving emission factor precision.

A key challenge in this work was the difficulty of obtaining data and information regarding the inputs and structure of Mobile5b. The effort required to do the uncertainty analysis once the data were available and the model was specified was a relatively small portion of this work. Uncertainty analysis is more efficient as an integral part of model development, rather than when done post hoc. Thus, we strongly support the NRC recommendation that uncertainty analysis should be an integral part of future emission factor models.

Acknowledgments

This work was supported in parts by the Center for Transportation and the Environment (CTE) at North Carolina State University and by the Office of Air Quality Planning and Standards of the U.S. Environmental Protection Agency (EPA). This paper has not been subject to any CTE or EPA review. Therefore, it does not necessarily reflect the views of CTE or EPA and no official endorsement should be inferred.

Supporting Information Available

Text, tables, and figures pertaining to Mobile5b input assumptions, model results, and verification of uncertainty analysis accuracy. The SI also includes analysis of uncertainty in Mobile6 speed correction factor driving cycles. This material is available free of charge via the Internet at http://pubs.acs.org.

Literature Cited

(1) National Research Council. *Modeling Mobile-Source Emissions;* National Academy Press: Washington, DC, 2000.
(2) TRB. *Expanding Metropolitan Highways: Implications for Air Quality and Energy Use*; Special Report 245; Transportation Research Board: Washington, DC, 1995.
(3) Kini, M. D.; Frey, H. C. *Probabilistic Evaluation of Mobile Source Air Pollution, Volume 1: Probabilistic Modeling of Exhaust Emissions from Light Duty Gasoline Vehicles*; Center for Transportation and the Environment, North Carolina State University: Raleigh, NC, 1997.
(4) Pollack, A. K.; Bhave P.; Heiken J.; Lee K.; Shepard S.; Tran C.; Yarwood G.; Sawyer R. F.; Joy B. A. *Investigation of Emission Factors in the California EMFAC7G Model*; PB99-149718INZ.; Coordinating Research Council: Atlanta, GA, 1999.
(5) Chatterjee, A.; Wholley, T. F., Jr.; Guensler, R.; et al. *Improving Transportation Data for Mobile Source Emissions Estimates*; Project 25−7; National Cooperative Highway Research Program: Washington, DC, 1996.
(6) *User's Guide to Mobile5*; EPA-AA-TEB-94-01; U.S. Environmental Protection Agency: Ann Arbor, MI. 1994 (Chapter 2 updated 1996).
(7) Heirigs P. L.; Dulla, R. G. *Investigation of Mobile5a Emission Factors: Evaluation of IM240-to-FTP Correlation and Base Emission Rate Equations*; API Publication 4605; American Petroleum Institute: Washington, DC, 1994.
(8) Sierra Research, Inc. *Evaluation of Mobile Vehicle Emission Model*; Prime Contract DTRS-57-8-D-00089; U.S. Department of Transportation: Washington, DC, 1994.
(9) Systems Applications International. *Investigation of Mobile5a Emission Factors*; Final Report SYSAPP94-93/21 lrl; American Petroleum Institute: Washington, DC, 1994.
(10) *Draft User's Guide to Mobile6: Mobile Source Emission Factor Model*; EPA420-D-01-003; U.S. Environmental Protection Agency: Ann Arbor, MI, 2001.
(11) Beardsley, M. *Mobile6: EPA's Highway Vehicle Emissions Model*; Presented at North American Vehicle Emission Control Conference; Atlanta, GA. 2001.
(12) Bogen, K. T.; Spear, R. C. *Risk Anal.* **1987**, *7*, 427−436.
(13) Efron, B.; Tibshirani, R. J. *An Introduction to the Bootstrap*; Chapman & Hall: New York, 1993.
(14) Hoffman, F. O.; Hammonds, J. S. *Risk Anal.* **1994**, *7*, 707−712.
(15) Frey, H. C.; Rhodes, D. S. *Hum. Health Ecol. Risk Assess.* **1996**, 2, 762−797.
(16) Burmaster, D. E.; Wilson, A. M. *Hum. Health Ecol. Risk Assess.* **1996**, 2, 892−919.
(17) Cullen, A.; Frey, H. C. *Probabilistic Techniques in Exposure Assessment*; Plenum: New York, 1999.
(18) Helton, J. C.; Bean, J. E.; Economy, K.; et al. *Reliablity Eng. Syst. Saf.* **2000**, *69*, 263−304.
(19) Frey, H. C.; Bharvirkar, R.; Zheng, J. *Quantitative Analysis of Variability and Uncertainty in Emissions Estimation*; U.S. Environmental Protection Agency, Research Triangle Park, NC, 1999.
(20) Webster, W. J.; Shih, C. A Statistically-Derived Metric to Monitor Time-Speed Variability in Practical Emissions Testing. *Proceedings of the Sixth CRC On-Road Vehicle Emissions Workshop*, Coordinating Research Council: Atlanta, GA, 1996.
(21) Zhang, Y.; Bishop, G. A.; Stedman, D. H. *Environ. Sci. Technol.* **1994**, *28*, 1370−1374.
(22) Bishop, G. A.; Stedman, D. H.; Ashbaugh, L. *J. Air Waste Manage. Assoc.* **1996**, *46*, 667−675.
(23) Frey, H. C.; Eichenberger, D. A. *Remote Sensing of Mobile Source Air Pollutant Emissions: Variability and Uncertainty in On-Road Emissions Estimates of Carbon Monoxide and Hydrocarbons for School and Transit Buses*, FHWY/NC/97-005, North Carolina Department of Transportation: Raleigh, NC, 1997.
(24) Brzezinski, D.; Hart, C.; Enns, P. Final Facility Specific Speed Correction Factors, EPA-420-R-01-060, U.S. Environmental Protection Agency: Ann Arbor, MI, 2001.

Received for review November 19, 2001. Revised manuscript received August 30, 2002. Accepted September 5, 2002.

ES0114308